Durham 15 May 2008

Decision Theory under Complex Uncertainty

Robert Hable
Department of Mathematics
University of Bayreuth

Durham 15 May 2008

Decision Theory under Complex Uncertainty

Robert Hable
Department of Mathematics
University of Bayreuth

$\longrightarrow$ but still Thomas's Ph.D. student

Decision Theory

Decision theory = Theory of “How to make good decisions, if this is necessary”

Decision Theory

Decision theory = Theory of "How to make good decisions, if this is necessary"

- "if this is necessary" → decision problem
- "decisions" → What can be done?
- "good decisions" → Optimality criterion?

Decision Theory

Decision theory = Theory of "How to make good decisions, if this is necessary"

- "if this is necessary" → decision problem
- "decisions" → What can be done?
- "good decisions" → Optimality criterion?

Examples:

- making an investment in a certain company
- Hypothesis testing: Reject the null hypothesis H_0 or not?
  → statistical decision theory

In the following: mathematical formalization of such things

Decisions and states of nature

Decisions

What are the possible decisions/actions?

$$\mathbb{D} \;=\; \text{set of all possible decisions } t$$

States of nature

- complete knowledge about the situation

  $$\text{situation} \quad = \quad \text{state of nature } \theta_0$$

  $\rightarrow$ decision making is very easy
- whole set of possible states of nature

  $$\theta \in \Theta$$

  Which one is true? $\quad \rightarrow \quad$ This is unknown!

Decisions and states of nature

Decisions

What are the possible decisions/actions?

$$\mathbb{D} \;=\; \text{set of all possible decisions } t$$

States of nature

- complete knowledge about the situation

  situation = state of nature θ_0

  $\longrightarrow$ decision making is very easy
- whole set of possible states of nature

$$\theta \in \Theta$$

Which one is true? $\longrightarrow$ This is unknown!

Will my favorite decision $t \in \mathbb{D}$ be a good one?

$\longrightarrow$ This depends on the unknown state of nature $\theta \in \Theta$!

Loss function

What are the consequences of my decision $t \in \mathbb{D}$?

Loss function

$$W : \quad \Theta \times \mathbb{D} \ \rightarrow \ \mathbb{R}\,, \quad (\theta, t) \ \mapsto \ W_\theta(t)$$

Every decision $t \in \mathbb{D}$ leads to a certain loss

$$W_\theta(t) \ \in \ \mathbb{R}$$

depending on the (unknown) state of nature $\theta \in \Theta$

Loss function

What are the consequences of my decision $t \in \mathbb{D}$?

Loss function

$$W: \quad \Theta \times \mathbb{D} \;\to\; \mathbb{R}\,, \quad (\theta, t) \;\mapsto\; W_\theta(t)$$

Every decision $t \in \mathbb{D}$ leads to a certain loss

$$W_\theta(t) \;\in\; \mathbb{R}$$

depending on the (unknown) state of nature $\theta \in \Theta$

Simple Example: Investment

- Decision t: Investment of 10 000 Euro in a company today
- State of nature θ: the company will collapse tomorrow
- loss $W_\theta(t)$: 10 000

Prior distribution

Probabilistic uncertainty

- We do not know which state of nature θ will occur!
- But: Sometimes we might know a precise prevision / probability measure for θ, i.e. a **precise prior distribution**

$$\pi : \ \mathcal{L}_\infty(\Theta) \to \mathbb{R}, \qquad h \mapsto \pi[h] = \int h(\theta)\, \pi(d\theta)$$

Prior distribution

Probabilistic uncertainty

- We do not know which state of nature θ will occur!
- But: Sometimes we might know a precise prevision / probability measure for θ, i.e. a **precise prior distribution**

$$\pi : \ \mathcal{L}_\infty(\Theta) \ \rightarrow \ \mathbb{R}\,, \qquad h \ \mapsto \ \pi[h] \ = \ \int h(\theta)\,\pi(d\theta)$$

Complex uncertainty

- We do not know which state of nature θ will occur!
- Knowing a precise prior distribution is totally unrealistic!
- But: We know an imprecise prevision for θ:

$$\overline{\pi} : \ \mathcal{L}_\infty(\Theta) \ \rightarrow \ \mathbb{R}\,, \qquad h \ \mapsto \ \overline{\pi}[h]$$

$\rightarrow$ This can be used to find a “good” decision.

What is a good decision?

Problem:

- It is hard to compare two decisions t_1 and t_2 because:

  Usually

  $$W_\theta(t_1) \; < \; W_\theta(t_2) \qquad \text{for some } \theta$$

  and

  $$W_{\theta'}(t_1) \; > \; W_{\theta'}(t_2) \qquad \text{for another } \theta'$$

  But: We do not know if θ or θ' will occur!

What is a good decision?

Problem:

- It is hard to compare two decisions t_1 and t_2 because:

  Usually
  $$W_\theta(t_1) \; < \; W_\theta(t_2) \qquad \text{for some } \theta$$
  and
  $$W_{\theta'}(t_1) \; > \; W_{\theta'}(t_2) \qquad \text{for another } \theta'$$
  But: We do not know if θ or θ' will occur!

What can be done?

- We have a prior distribution / prevision for θ !
  $\longrightarrow$ Deciding according to the expected loss

What is a good decision?

Precise prior distribution

- Find a decision $\tilde{t}$ which minimizes

$$\pi[W_{\cdot}(t)] \;=\; \int_{\Theta} W_{\theta}(t)\, \pi(d\theta) \;\in\; \mathbb{R}$$

What is a good decision?

Precise prior distribution

- Find a decision $\tilde{t}$ which minimizes

$$\pi[W.(t)] \;=\; \int_{\Theta} W_\theta(t)\, \pi(d\theta) \;\in\; \mathbb{R}$$

Imprecise prior distribution

- The imprecise prior distribution leads to a whole interval

$$\big[\underline{\pi}[W.(t)]\,,\,\overline{\pi}[W.(t)]\big] \;\subset\; \mathbb{R}$$

of expected losses.

$\longrightarrow$ How to compare overlapping intervals???

$\longrightarrow$ Different possibilities!

Different optimality criteria

- The imprecise prior distribution leads to a whole interval

$$\left[\underline{\pi}[W_{\cdot}(t)]\,,\,\overline{\pi}[W_{\cdot}(t)]\right] \;\subset\; \mathbb{R}$$

of expected losses.

$\longrightarrow$ How to compare overlapping intervals?

$\longrightarrow$ What is a good decision $t \in \mathbb{D}$?

Different optimality criteria

- The imprecise prior distribution leads to a whole interval

$$\left[\underline{\pi}[W.(t)]\,,\,\overline{\pi}[W.(t)]\right] \;\subset\; \mathbb{R}$$

of expected losses.

$\longrightarrow$ How to compare overlapping intervals?

$\longrightarrow$ What is a good decision $t \in \mathbb{D}$?

- Different optimality criteria:
  - Γ-minimax, Γ-minimin and a mixture
  - E-admissibility
  - maximality
  - interval dominance

Γ-minimax, Γ-minimin and a mixture

- Γ-**minimax**: Find a decision $\tilde{t}_1$ which minimizes the upper bound

$$\overline{\pi}[W.(t)] \ = \ \text{upper expected loss}$$

$\longrightarrow$ $\tilde{t}_1$ is optimal in the worst case (pessimistic)!

Γ-minimax, Γ-minimin and a mixture

- Γ-**minimax**: Find a decision $\tilde{t}_1$ which minimizes the upper bound

$$\overline{\pi}[W.(t)] \;=\; \text{upper expected loss}$$

$\longrightarrow$ $\tilde{t}_1$ is optimal in the worst case (pessimistic)!

Justification: Murphy's law

Γ-minimax, Γ-minimin and a mixture

- Γ-**minimax**: Find a decision $\tilde{t}_1$ which minimizes the upper bound

$$\overline{\pi}[W.(t)] \;=\; \text{upper expected loss}$$

$\rightarrow$ $\tilde{t}_1$ is optimal in the worst case (pessimistic)!

Γ-minimax, Γ-minimin and a mixture

- Γ-**minimax**: Find a decision $\tilde{t}_1$ which minimizes the upper bound

$$\overline{\pi}[W_{\cdot}(t)] \;=\; \text{upper expected loss}$$

$\rightarrow$ $\tilde{t}_1$ is optimal in the worst case (pessimistic)!

- Γ-**minimin**: Find a decision $\tilde{t}_2$ which minimizes the lower bound

$$\underline{\pi}[W_{\cdot}(t)] \;=\; \text{lower expected loss}$$

$\rightarrow$ $\tilde{t}_2$ is optimal in the best case (optimistic)!

Γ-minimax, Γ-minimin and a mixture

- **Γ-minimax**: Find a decision $\tilde{t}_1$ which minimizes the upper bound

$$\overline{\pi}[W.(t)] \;=\; \text{upper expected loss}$$

$\rightarrow$ $\tilde{t}_1$ is optimal in the worst case (pessimistic)!

- **Γ-minimin**: Find a decision $\tilde{t}_2$ which minimizes the lower bound

$$\underline{\pi}[W.(t)] \;=\; \text{lower expected loss}$$

$\rightarrow$ $\tilde{t}_2$ is optimal in the best case (optimistic)!

- **and a mixture**: Find a decision $\tilde{t}_3$ which minimizes a convex combination of the lower and the upper bound

$$\alpha\overline{\pi}[W.(t)] + (1-\alpha)\underline{\pi}[W.(t)]$$

$\rightarrow$ sensible(?) compromise between pessimism and optimism
α : caution parameter

E-admissibility and Maximality

Let

$$\mathcal{P} \;=\; \left\{\pi \in \mathcal{M}_1(\Theta) \;\middle|\; \underline{\pi}[h] \leq \pi[h] \leq \overline{\pi}[h] \;\; \forall\, h \in \mathcal{L}_\infty(\Theta)\right\}$$

be the credal set of the imprecise prevision.

E-admissibility

- Find a decision $\tilde{t}_4$ which minimizes

$$\pi[W_\cdot(t)] \;=\; \int_\Theta W_\theta(t)\,\pi(d\theta)$$

for at least one $\pi \in \mathcal{P}$.

E-admissibility and Maximality

Let

$$\mathcal{P} \;=\; \left\{\pi \in \mathcal{M}_1(\Theta) \;\middle|\; \underline{\pi}[h] \le \pi[h] \le \overline{\pi}[h] \;\; \forall\, h \in \mathcal{L}_\infty(\Theta)\right\}$$

be the credal set of the imprecise prevision.

E-admissibility

- Find a decision $\tilde{t}_4$ which minimizes

$$\pi[W_\cdot(t)] \;=\; \int_\Theta W_\theta(t)\,\pi(d\theta)$$

for at least one $\pi \in \mathcal{P}$.

Maximality

- Find a decision $\tilde{t}_5$ such that: For every $t \in \mathbb{D}$, there is at least one $\pi_t \in \mathcal{P}$ such that

$$\pi_t[W_\cdot(\tilde{t}_5)] \;\le\; \pi_t[W_\cdot(t)]$$

Interval dominance

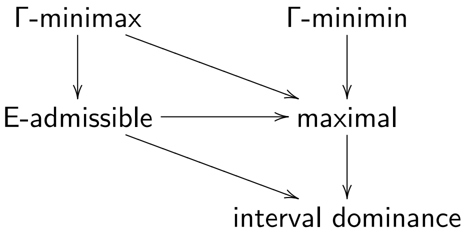
- A decision $\tilde{t}_6$ is optimal with respect to **interval dominance** if

  there is no decision $t \in \mathbb{D}$ such that

$$\overline{\pi}\big[W_{\cdot}(t)\big] \;\leq\; \underline{\pi}\big[W_{\cdot}(\tilde{t}_6)\big]$$

Which is the right one?

The following implications hold:



Confer Troffaes (2007).

Data-based decision theory

Often: Decision making on base of observations is possible

$$\text{Observation / data } x \qquad \longrightarrow \qquad \text{decision } t(x)$$

Example: Statistical decisions should be based on data

Data-based decision theory

Often: Decision making on base of observations is possible

$$\text{Observation / data } x \quad \longrightarrow \quad \text{decision } t(x)$$

Example: Statistical decisions should be based on data

So, we have

- a **sample space** $(\mathcal{X}, \mathcal{A})$
- an **observation** $x \in \mathcal{X}$
- a **distribution of the observation**

$$x \sim P_\theta$$

which depends on the (unknown) state of nature.

Data-based decision theory

Decision making on base of observations

$$\text{Observation} \,/\, \text{data } x \qquad \longrightarrow \qquad \text{decision } t(x)$$

Data-based decision theory

Decision making on base of observations

$$\text{Observation / data } x \qquad \longrightarrow \qquad \text{decision } t(x)$$

So,

- we have to chose not only one decision t but a decision function

$$\delta: \ \mathcal{X} \to \mathbb{D}\,, \qquad x \mapsto \delta(x)$$

- or even a randomized decision function, i.e. a Markov kernel

$$\tau: \ \mathcal{X} \times \mathcal{D} \to \mathbb{R}\,, \qquad (x, D) \mapsto \tau_x(D)$$

(τ_x a probability measure on $(\mathbb{D}, \mathcal{D})$)

Data-based decision theory

The expected loss of a decision function $\delta : x \mapsto \delta(x)$

- **Precise previsions**:

$$\int_{\Theta} \int_{\mathcal{X}} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta)$$

Data-based decision theory

The expected loss of a decision function $\delta : x \mapsto \delta(x)$

- **Precise previsions**:

$$\int_\Theta \int_\mathcal{X} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta)$$

- **Imprecise previsions**:

$$\left\{ \int_\Theta \int_\mathcal{X} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta) \;\middle|\; \pi \in \mathcal{P}\,,\;\; P_\theta \in \mathcal{M}_\theta \;\; \forall\, \theta \in \Theta \right\}$$

  - $\mathcal{P}$ credal set of an imprecise prevision of θ
  - $\mathcal{M}_\theta$ credal set of an imprecise prevision of the observation x
    $\rightarrow$ imprecise model $\big(\mathcal{X}, \mathcal{A}, (\mathcal{M}_\theta)_{\theta\in\Theta}\big)$

Data-based decision theory

In case of imprecise previsions:

- What is an optimal decision function?

Data-based decision theory

In case of imprecise previsions:

- What is an optimal decision function?
- e.g. Γ-**minimax** criterion

$$\sup \left\{ \int_\Theta \int_\mathcal{X} W_\theta(\delta(x)) \, P_\theta(dx) \, \pi(d\theta) \;\middle|\; \ldots \right\} =$$

$$= \sup_{\pi \in \mathcal{P}} \int_\Theta \sup_{P_\theta \in \mathcal{M}_\theta} \int_\mathcal{X} W_\theta(\delta(x)) \, P_\theta(dx) \, \pi(d\theta)$$

Data-based decision theory

In case of imprecise previsions:

- What is an optimal decision function?
- e.g. Γ-**minimax** criterion

$$\sup \left\{ \int_\Theta \int_\mathcal{X} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta) \,\middle|\, \ldots \right\} =$$

$$= \sup_{\pi \in \mathcal{P}} \int_\Theta \sup_{P_\theta \in \mathcal{M}_\theta} \int_\mathcal{X} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta)$$

→ widely accepted optimality criterion in robust statistics

Data-based decision theory

In case of imprecise previsions:

- What is an optimal decision function?
- e.g. Γ-**minimax** criterion

$$\sup \left\{ \int_\Theta \int_\mathcal{X} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta) \;\Big|\; \ldots \right\} =$$

$$= \sup_{\pi \in \mathcal{P}} \int_\Theta \sup_{P_\theta \in \mathcal{M}_\theta} \int_\mathcal{X} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta)$$

$\rightarrow$ widely accepted optimality criterion in robust statistics

- How to find the optimal decision function?

Data-based decision theory

In case of imprecise previsions:

- What is an optimal decision function?
- e.g. Γ-**minimax** criterion

$$\sup \left\{ \int_{\Theta} \int_{\mathcal{X}} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta) \;\middle|\; \ldots \right\} =$$

$$= \sup_{\pi \in \mathcal{P}} \int_{\Theta} \sup_{P_\theta \in \mathcal{M}_\theta} \int_{\mathcal{X}} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta)$$

$\longrightarrow$ widely accepted optimality criterion in robust statistics

- How to find the optimal decision function?
  $\longrightarrow$ This is a big problem, in general!

Data-based decision theory

Do we need to consider data-based decision theory?

Data-based decision theory

Do we need to consider data-based decision theory?

- **Precise previsions**: No!
  $\longrightarrow$ main theorem of Bayesian decision theory

Data-based decision theory

Do we need to consider data-based decision theory?

- **Precise previsions**: No!
  $\longrightarrow$ main theorem of Bayesian decision theory
- **Imprecise previsions**: Yes

Data-based decision theory

Do we need to consider data-based decision theory?

- **Precise previsions**: No!
  $\longrightarrow$ main theorem of Bayesian decision theory
- **Imprecise previsions**: Yes, I think so!
  $\longrightarrow$ Confer Augustin(2003) and Thomas's talk later on.

References

- T. Augustin. On the suboptimality of the generalized Bayes rule and robust Bayesian procedures from the decision theoretic point of view: a cautionary note on updating imprecise priors. In J.M. Bernard, T. Seidenfeld, and M. Zaffalon, editors, *ISIPTA'03, Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications, Lugano*, pages 31-45. Carleton Scientifc, Waterloo, 2003.
- J.O. Berger. *Statistical decision theory and Bayesian analysis.* Springer-Verlag, New York, second edition, 1985.
- R. Hable. *Data-based decisions under complex uncertainty.* Ph.D. Thesis, LMU Munich, in preparation.
- M.C.M. Troffaes. Decision making under uncertainty using imprecise probabilities. *International Journal of Approximate Reasoning*, 45:17-29, 2007.